Large-Scale Visual Search

Qi Tian

Department of Computer Science
University of Texas at San Antonio, USA
qitian@cs.utsa.edu

11/25/2013 UTSA
Outline

• Background
  – Overview
  – 1st Generation CBIR
  – 2nd Generation CBIR

• Key Problems

• Our Work

• Trends & Discussions
Popularity of digital camera

Huge web video data
72 hours/minute in April 2012

50 Billion photos, 2.5 billion/month in July 2010

Explosive Growth of Multimedia Data

Large-scale Visual Search

- Text based Image Search
  - Labor extensive
  - Human subjective

- Content based Image Search (CBIR)
  - color, texture, shape
The growing list: ADL, AltaVista Photofinder, Amore, ASSERT, BDLP, Blobworld, CANDID, C-bird, Chabot, CBVQ, DrawSearch, Excalibur Visual RetrievalWare, FIDS, FIR, FOCUS, ImageFinder, ImageMiner, ImageRETRO, ImageRover, ImageScape, Jacob, LCPD, MARS, MetaSEEk, MIR, NETRA, Photobook, Picasso, PicHunter, PicToSeek, QBIC, Quicklook2, SIMBA, SQUID, Surfimage, SYNAPSE, TODAI, VIR Image Engine, VisualSEEk, VP Image Retrieval System, WebSEEk, WebSeer, WISE...
Image Search Applications

Web based Image Search

Mobile based Image Search

CD cover
Logo
Artwork
Landmark
tattoo
pedestrian
Search by Mobile
Goal of CBIR

- **Database:** millions -> billions
  - Retrieve identical images (√)
  - Retrieve nearly duplicated images: cropped, edited, resized (√)
  - Retrieve images containing similar objects (√)
  - Retrieve similar semantics (X)

Easy vs. Difficult
Challenge I - Semantic Gap

- Similar visual content (color, texture, shape)
- Different semantics

**Semantic Gap:** the gap between low-level visual features and high-level concepts
Challenge - Semantic Gap
Challenge II – Intention Gap

Intention Gap – the gap between user search intent and the query

• Users can imagine what they want, but cannot express it in precise words.
Outline

• Background
  – Overview
  – 1\textsuperscript{st} Generation CBIR
  – 2\textsuperscript{nd} Generation CBIR

• Key Problems

• Our Work

• Trends & Discussions
1st-Generation CBIR (early 90’s – 2003)

General Framework

Image DB (<100k)

Index Builder

Feature Extraction

Image Ranking

Query

Query Formulation

Feature Extraction

Similarity Computation

Relevance Feedback

Re-ranking

Web

Image Crawler

Offline

Online
Query Schemes

Query by keyword
- dog
- dog
- dog

Query by example
- couple
- couple
- couple
- couple
- couple
- couple
- couple
- couple

Query by sketch
- sketch
- sketch
- sketch

Query by color layout
- jeep
- grass

Query by concept layout
- jeep
- grass

Abstract Thoughts

Interface

Global Feature → Similarity → Index → Relevance Feedback → Re-ranking
**Global Feature**

- Low-level visual representation from the whole image

- Feature vector
  - Color moments
  - Color histogram
  - Color correlogram
  - Fourier descriptor
  - Shape context
  - Tamura feature
  - Gabor feature
  - Co-occurrence matrix

---

**Query**

**Global Feature**

**Similarity**

**Index**

**Relevance Feedback**

**Re-ranking**
Global Feature

Images are represented by feature vectors.
Problems with Global Feature

Successful results

How to bridge the Semantic Gap?
- Extract discriminative features
- Learn discriminative distance metric

Color Feature Extraction

Semantic Gap

Failed results
Distance Metric Learning

Two types:
- **Supervised** distance metric learning
  - *Mahalanobis* Distance
- **Unsupervised** distance metric learning
  - Find an embedding in a lower dimensional space
  - Also called manifold learning
    - PCA, MDS, ISOMAP, LLE, Laplacian Eigenmap

**Distance Metric Learning**

- Similar semantics $\rightarrow$ smaller distance
- Dissimilar semantics $\rightarrow$ larger distance
Fast Similar Image Search

- **Alternative**
  - Find approximate nearest neighbors, with some sacrifice in accuracy
  - Avoid exhaustive search (\(O(n)\) time complexity), ensure efficiency

- **Existing methods**
  - KD-Tree, LSH (Locality Sensitive Hashing)
Relevance Feedback

Human in the loop
- Collaborative vs. passive users

Types of feedback
- **Explicit feedback:**
  - Positive examples only
  - Positive and negative examples
- **Implicit feedback:** system infers user intentions based on observable behavior
- **Blind feedback:** pseudo feedback

Rui, et al. TCSVT 1998
Image Search Re-ranking

Text Query: “beach”

Text-based search

Initial ranked list

Laguna Beach wave: Canon EOS 1Ds Mark

Caribbean Soul Navarre Beach (Panhandle)...

Hotels at beach in Dominican Republic...

This beach offers 3 miles of undeveloped...

Beach deposits predominantly consist of sand...

Paradise restaurant terraza bar and beach club...

Visual information

Re-ranking

Re-ranked list
Outline

• Background
  – Overview
  – 1\textsuperscript{st} Generation CBIR
  – 2\textsuperscript{nd} Generation CBIR

• Key Problems

• Our Work

• Trends & Discussions
2\textsuperscript{nd} Generation CBIR (2003 – 2012)

A picture is worth a thousand words

But which 1000 words?
Of all the sensory impressions proceeding to the brain, the visual experiences are the dominant ones. Our perception of the world around us is based essentially on the messages that reach the brain from our eyes. For a long time it was thought that the retinal image was transmitted point by point to visual centers in the brain. The cerebral cortex was a movie screen, so to speak, upon which the image in the eye was projected.

"sensory, brain, visual, perception, retinal, cerebral cortex, eye, cell, optical nerve, image Hubel, Wiesel"

China is forecasting a trade surplus of $90bn (£51bn) to $100bn this year, a threefold increase on 2004's $32bn. The Commerce Ministry said the surplus would be created by a predicted 30% jump in exports to $750bn, compared with a 18% rise in imports to $660bn. The figures are likely to further annoy the US, which has long argued that China's exports are unfairly helped by a deliberately undervalued yuan.

"China, trade, surplus, commerce, exports, imports, US, yuan, bank, domestic, foreign, increase, trade, value"
What is a Visual Word?

- **Analogy to documents**
- Could images be represented as Bag-of-Visual Words?
  - Leverage classic methods in IR for image applications

Courtesy: Fei-Fei Li
General Framework

2\textsuperscript{nd} Generation CBIR

Query

Feature Extraction

Vector Quantization

Index Lookup

…… Retrieval Results

On-line

Database

Feature Extraction

Vector Quantization

Image Index

Off-line

Codebook Training

Feature Extraction

Vector Quantization

Image Index

…… Retrieval Results
Local Feature Extraction

- Local feature: representation of an image patch

**Question 1:** Where is the location of the patch?

**Interest Point/Region Detector**

**Question 2:** How to represent the patch into a vector?

**Patch Descriptor**
Local Feature Extraction

• **Interest Point Detector**
  – Detect interest points
    • **Stable** to image changes, e.g., zoom-in, zoom-out, rotation, affine transformations, illumination, etc.
    • **Highly repeatable** among visually similar images.

• **Local Patch Descriptor**
  – Describe the detected regions around interest points
    • **Efficient** to extract - Speed
    • **Compact** to store - Memory cost
    • **Discriminative** - Performance
    • **Robust** to scale, orientation, Affine transformation changes
Visual Vocabulary Generation

Feature Extraction → Codebook and Quantization → Indexing → Post Processing

Feature space

SIFT

Visual Word

Feature Clustering
K-means, Affinity Propagation, etc.
Feature Quantization

- From feature to visual word
  - Map a high-dimension vector to the closest visual word represented as an integer (visual word id)

- Popular techniques
  - Hierarchical 1-NN

Visual Word

KD-tree
Image Representation using Visual Word Histogram

Feature Extraction → Codebook and Quantization → Local Features Extraction → Local Features → Local Features Quantization → Visual word histogram → Bag of visual words → Post Processing

Image

Local Features

Visual word histogram

Courtesy: Fei-Fei Li
Inverted Indexing

- Most commonly used indexing strategy:

$$V_i$$

$$\text{IDF}(V_i)$$

Indexed feature

Indexed feature

Indexed feature

Indexed feature

Image ID of \(d\)

Term frequency \(tf_d(V_i)\)

Nister, Scalable Recognition with a Vocabulary Tree, *CVPR 06.*
Most commonly used ranking strategy:
- \(tf\) weighted similarity
- \(tf-IDF\) weighted similarity:

Nister, Scalable Recognition with a Vocabulary Tree, *CVPR 06.*
Query Expansion

• **Goal**: Enhance original query with more local features

• **Key ideas**:
  – Precision of original query shall be high.
  – Expanded query must focus on duplicated patch.

• Transitive closure expansion

Geometric Verification

- **Goal:** Remove false matches by checking geometric consistency

- **Popular method:** RANSAC
Outline

• Background
• **Key Problems**
• Our Work
• Trends & Discussions
Key Problems

- Computational Efficiency
- Codebook Training
- Image Indexing
- Spatial Context/Semantic Embedding
- Query Image
- Feature Extraction
- Feature Quantization
- Lookup Index
- Search Reranking
- Dynamic Data
- Quantization Loss
- Geometric Context Verification
- Retrieval Results
Local Feature Extraction:
- Binary SIFT (MM 2012)
- Edge-SIFT (TIP 2013)
- USB (TIP 2013, in revision)
Approaches to Key Problems

1. **Feature Extraction**
   - Query Image
2. **Feature Quantization**
   - Loss
   - Dynamic data
3. **Codebook Training**
   - Descriptive Visual Word and Descriptive Visual Phrase (TIP, 2011)
   - Scalable Cascaded Hashing (TMM, 2013)
4. **Image Indexing**
5. **Lookup Index**
6. **Search Reranking**
7. **Retrieval Results**
Approaches to Key Problems

Image Indexing
- Semantic-aware Co-indexing (ICCV 2013)
- Spatial Context Embedding (MM, 2012)

Spatial Context/ Semantics Embedding
Approaches to Key Problems

- Feature Extraction
- Query Image
- Retrieval Results
- Search
- Reranking
- Lookup Index
- Feature Quantization
- Image Indexing
- Codebook Training
- Spatial Coding (MM 2010)
- Geometric Coding (TOMCCAP 2013)

Geometric Context Representation
Outline

• Background
• Key Problems
• Our Work
• Trends & Discussions
Our Work

- **Local Feature Extraction**
  - Developed a set of binary feature descriptors
    - Fast, effective and memory efficient

- **Codebook Construction and Quantization**
  - Construct a discriminant and descriptive visual codebook
  - Construct a training-free visual codebook
    - No training cost, easy to implement, very effective, scalable

- **Image Indexing**
  - Semantic-aware and Spatial Context Embedding
    - Fuse the similar image search (Semantic Attributes) and near-duplicate image search (BoW model)

- **Geometric Verification**
  - Effective to remove false SIFT matches
  - Very fast and scalable

- **Applications**
  - Vehicle License Plate Detection
  - Vehicle Logo Detection
  - Social-sense Image Search
  - Fine-Grained Image Classification (Birds, Dogs, Flowers, Food)
Our Work

– Local Feature Extraction
  • Binary SIFT – MM 2012
  • Edge-SIFT – TIP 2013
  • Ultra Short Binary Descriptor (USB) – TIP 2013 in revision

– Codebook and Quantization
  • Descriptive Visual Word and Descriptive Visual Phrase – TIP 2011
  • Contextual Visual Vocabulary – MM 2010 and CVIU 2011
  • Scalable Cascaded Hashing – TMM 2013

– Image Indexing
  • Semantic-aware Co-indexing – ICCV 2013
  • Spatial Context Embedding – MM 2012

– Geometric Verification
  • Spatial Coding – MM 2010
  • Geometric coding – TOMCCAP 2013

– Applications
  • Vehicle License Plate Detection - TIP 2012
  • Vehicle Logo Detection and Recognition
  • Social-sense Image Search – CVIU 2013
  • Fine-Grained Image Classification – ICCV 2013
Real time retrieval in the database containing 2 Million images
Image Search Demos

• Spatial Coding for Large-scale Image Search (MM 2010)
  – 1 million image
  – Video on 10 million image search

• Codebook Training-Free Visual Search (MM 2012)
  – 1 million image

• Semantic-aware Co-indexing (ICCV 2013)
  – 1.3 million images

• USB Feature Matching and Image Retrieval (TIP 2013 in revision)
  – 1.3 million images

• License Plate Detection (TIP 2012)
Outline

• Background
• Key Problems
• Our Work
• Trends & Discussions
Perspective

1st generation CBIR: Global Feature Based

2nd generation CBIR: Codebook training Based

3rd generation CBIR:
- Binary local descriptor
- Codebook training-free quantization
- Co-indexing structure
- Web image mining

Year

Data scale

1K 10K 10M 1B
Codebook Training-Free with Binary Feature

Database

[0 1 0 ...... 1 0]  
[1 1 0 ...... 0 0]  
Binary Feature Extraction

Image Index
General Co-indexing Structure

Image DB

Visual Features

Index visual features

Co-indexing

Web

Metadata, Social network user comments, etc.

Knowledge mining

• User Preference
• Popularity
• GPS info.
• Image Category
Image Mining

Object Representation

Social mining

Image wiki:
Celebrity image for mining

Meta association

- Tags
- Surrounding texts
- Time, Place: GPS to virtual earth
- Image history

Recommendation

- Content-based image suggestion
- Cross media
  - Image to video search
  - Image snippet to webpage

Image Mash-up

- Enhancement & Super-resolution
- Scene completion
  2D images to 3D

Image Search and Re-Search

Meta data

Start with Images...
Long Lasting Battle to Bridge the Semantic Gap

**Concept**

**Giraffe**

**Semantic Gap**

- **Tiger, Creek, Sky, Grass, Dog**
  - **ObjectNet**
  - **Web image mining**

- **DVW, DVP, Bundled Feature, Contextual Visual Vocabulary**
  - **Refined**

- **Local Features**
  - SIFT, Salient Points, Visual Word, Image Patches
  - **Noisy**

- **Regional Features**
  - Region-of-Interests, Segmentation, Multiple Instances
  - **Difficult**

- **Global Features**
  - Color Histogram, Texture, Color Correlogram, edge-map
  - **Coarse**
Acknowledgement
Acknowledgement

Wengang Zhou, USTC

Shiliang Zhang, UTSA

Rongrong Ji, Xiamen University

Xinmei Tian, USTC

Guojun Qi, UIUC

Zheng-Jun Zha, CAS
Questions?

Thanks!